



Identificação e sequenciação de genes da via da lenhificação em bibliotecas BAC de *Eucalyptus globulus*

Ricardo Joel Martins Bernardino Marques Barrela

Dissertação para a obtenção de grau de Mestre em
Engenharia Florestal e dos Recursos Naturais

Orientador: Jorge Almiro Barceló Caldeira Pinto Paiva
Co-orientador: Maria Helena de Noronha Ribeiro de Almeida

Lisboa, 12 de Dezembro 2011

Introdução

Eucalyptus globulus

- Espécie preferencial para a produção de pasta nas regiões temperadas;
- Em Portugal a espécie ocupa 20,6% ($646,7 \times 10^3$ ha);
- Indústria da pasta e do papel é responsável por 3% das exportações nacionais;
- Portugal é o 3º maior produtor europeu de pasta.



Introdução

Composição química da madeira de Eucalipto

	Celulose (%)	Hemicelulose (%)	Lenhina total (%)	S/G (mol/mol)
<i>E. globulus</i>	48,8	24,5	26,7	4,3
<i>E. nitens</i>	43,8	30,9	25,3	2,7
<i>E. grandis</i>	51,7	20,5	27,8	3,2
<i>E. urograndis</i>	47,5	24,5	28,0	2,0
<i>E. urophyla</i>	52,0	19,4	28,6	2,4

Fonte: Magaton et al. 2006

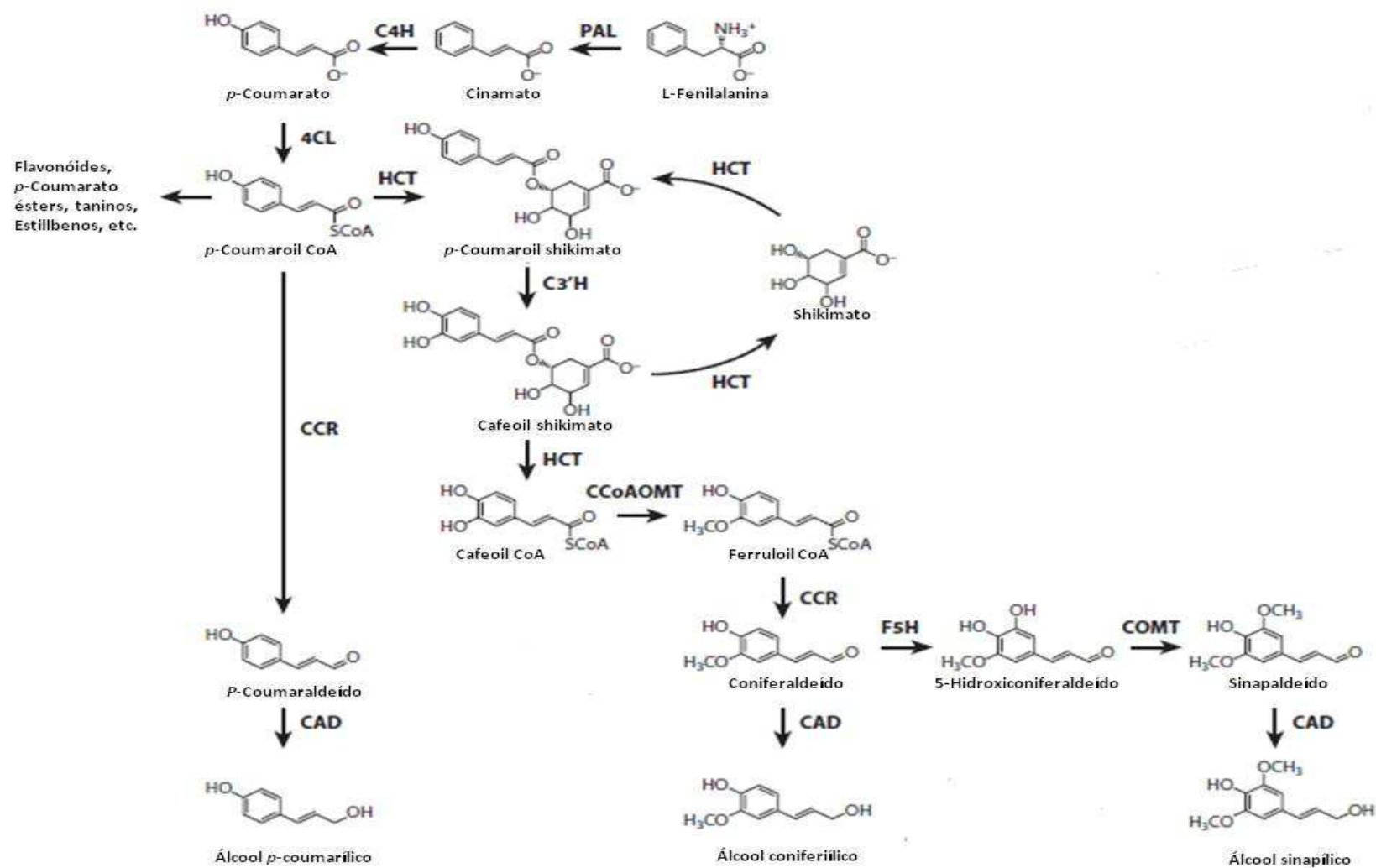
Composição química influi no rendimento/qualidade da pasta



Controlo da qualidade e quantidade da lenhina

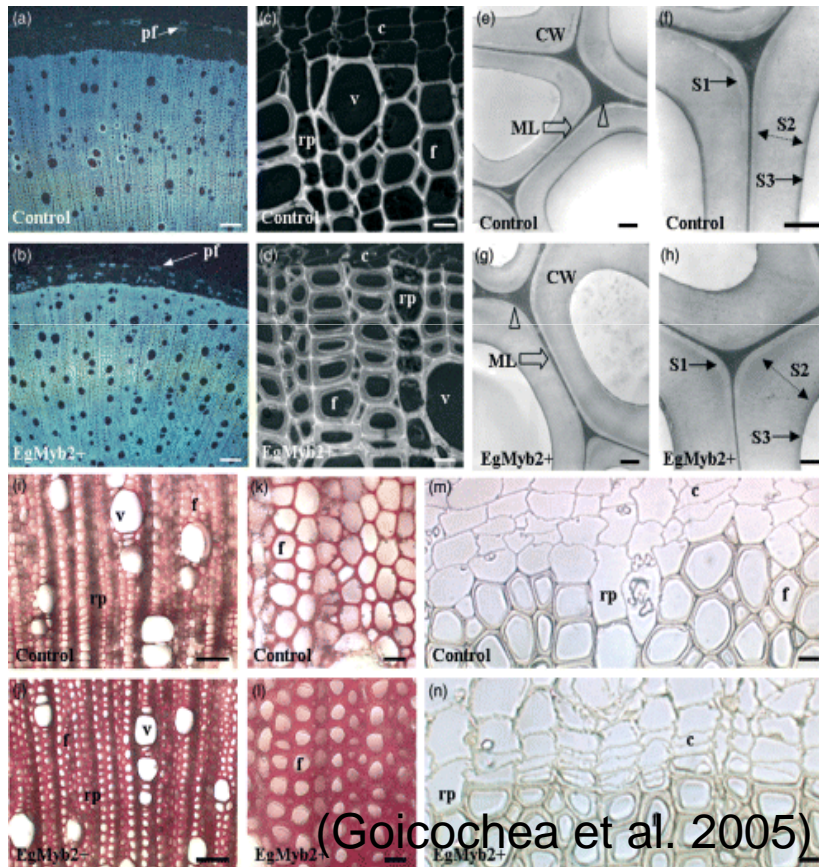
Introdução

Via biossintética das lenhinas



Introdução

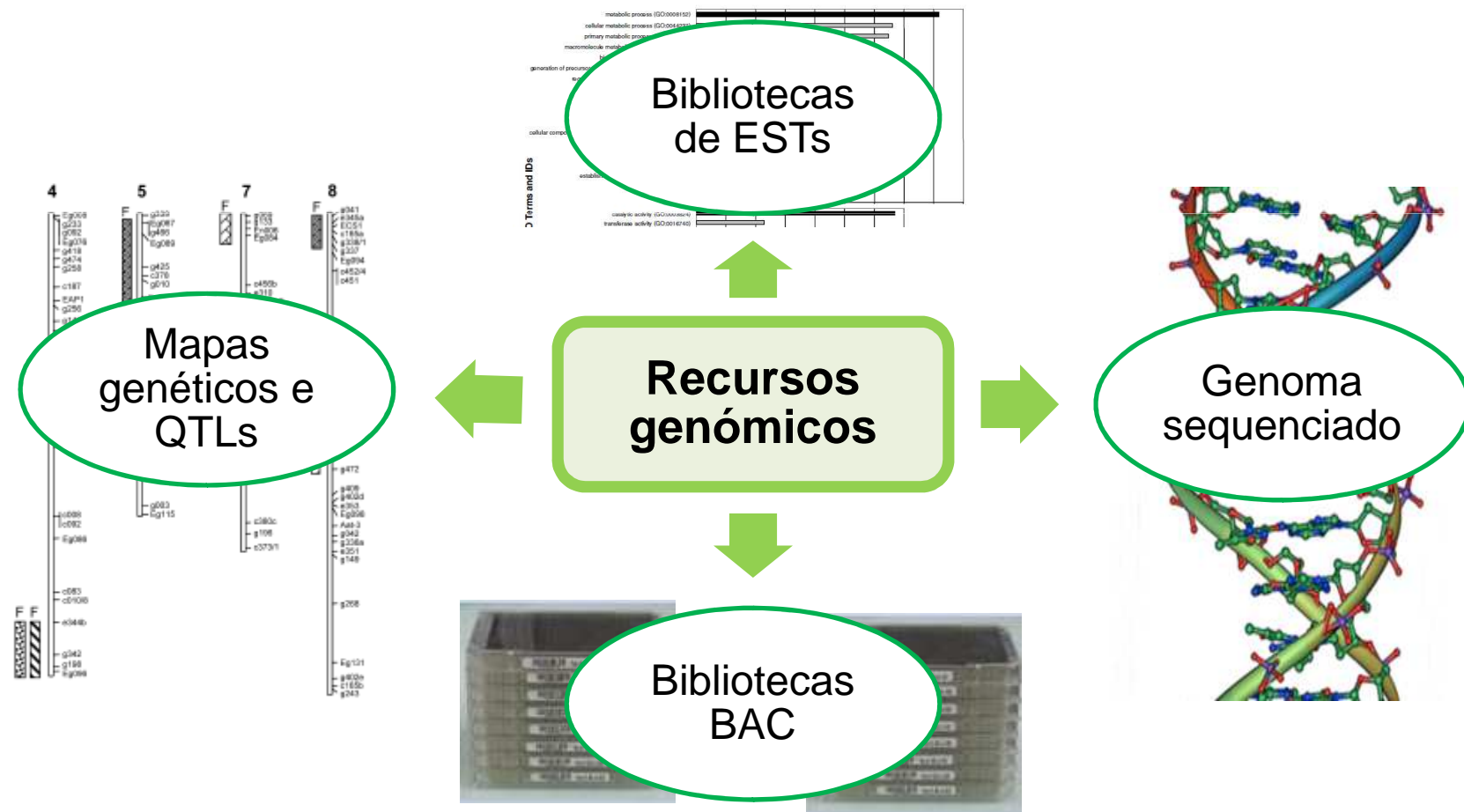
Regulação da biossíntese de lenhina



- *EgMyb1* (Legay et al. 2007);
- *EgMyb2* (Goicochea et al. 2005);
- *EgRac1* (Foucart et al. 2006).

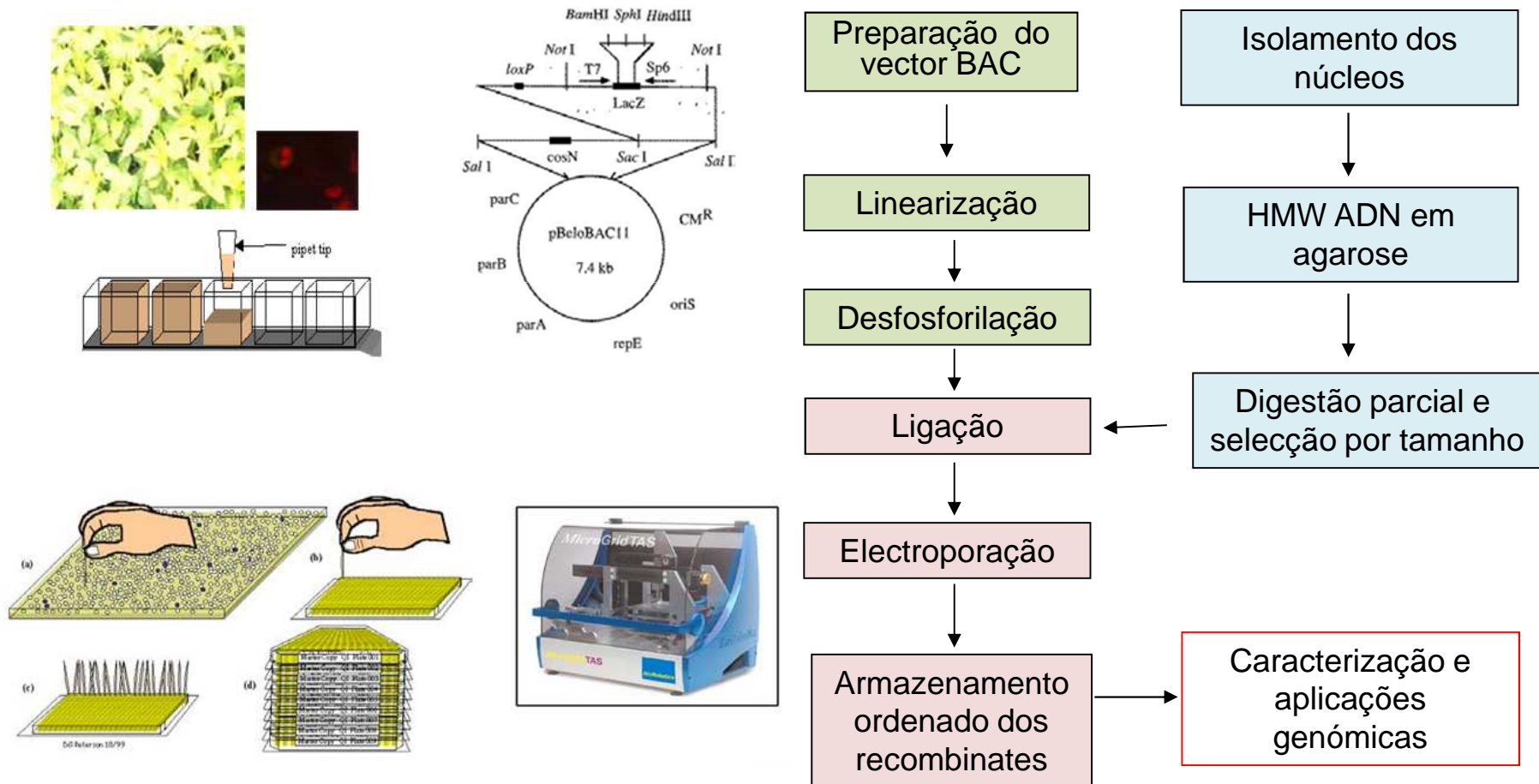
Introdução

Eucalipto enquanto espécie modelo



Introdução

Biblioteca BAC



Objectivos

Conhecer o número e estrutura dos genes intervenientes na produção e acumulação de lenhina em *E.globulus* e, caracterizar a região genómica envolvente.



1. Caracterizar bibliotecas BAC *E. globulus*
2. Identificar clones BAC portadores de genes de interesse
3. Caracterizar regiões do genoma nuclear.

1. Caracterização das bibliotecas

- **Bibliotecas de BAC *E. globulus***

Clemson University Genomics Institute (CUGI, Clemson University, USA; <http://www.genome.clemson.edu/>), RAIZ.

- ADN obtido a partir de folhas jovens e ensombradas (clone PL131)
- ***Eco*RI (EGC_Ba) e *Hind*III (EGC_Bb)**
- 36.864 clones, 96 placas (384 poços)
- Tamanho médio dos insertos, 115kb
- Cobertura do genoma $\approx 16X$.

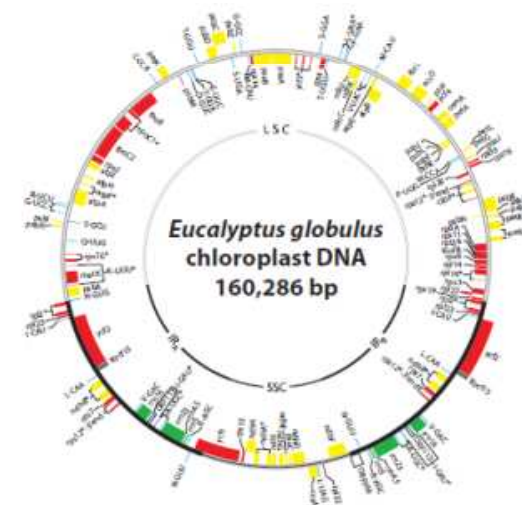


1. Caracterização das bibliotecas

Contaminação por genoma extra-nuclear

- **Hibridação de membranas** de colónias de alta densidade de BAC com sondas de interesse

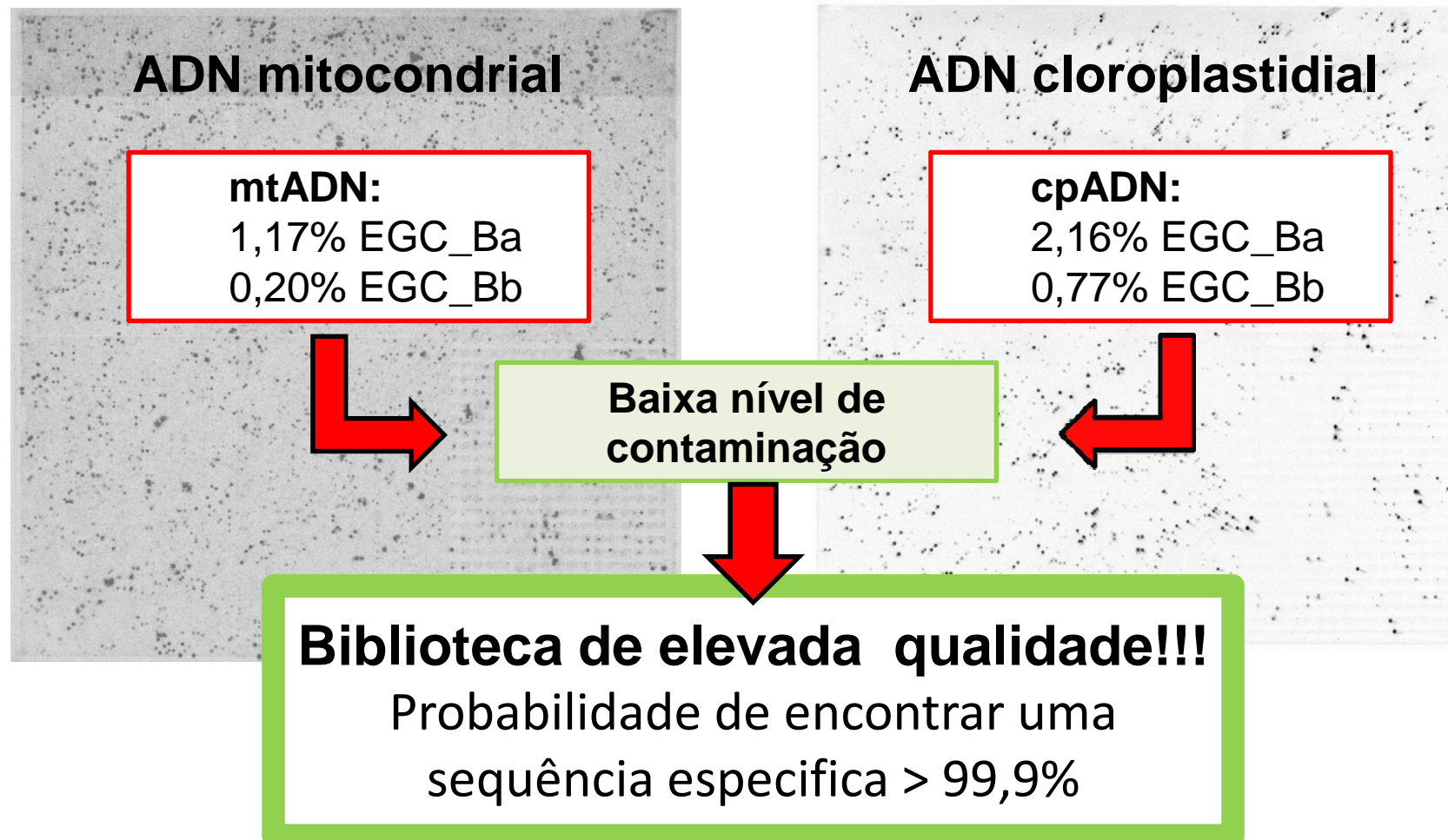
Genoma	Gene
Cloroplasto	<i>psbA</i>
	<i>psbB</i>
	<i>psbD</i>
	<i>rbcL</i>
	<i>ndhB</i>
Mitocôndria	<i>ccb256</i>
	<i>cox3</i>



- Reconfirmação dos clones positivos *via* PCR

1. Caracterização das bibliotecas

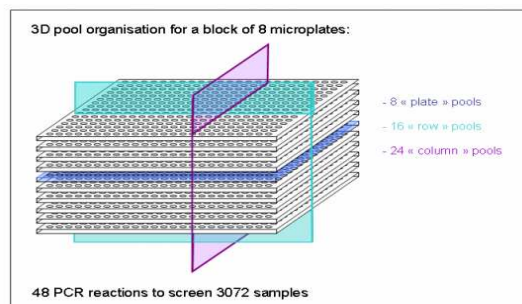
Resultados: Contaminação por genoma extra-nuclear



2. Identificação de clones BAC portadores de genes de interesse

Estratégia:

- a) Construção de *Pools* 3D das bibliotecas BAC
- b) Identificação de clones por PCR

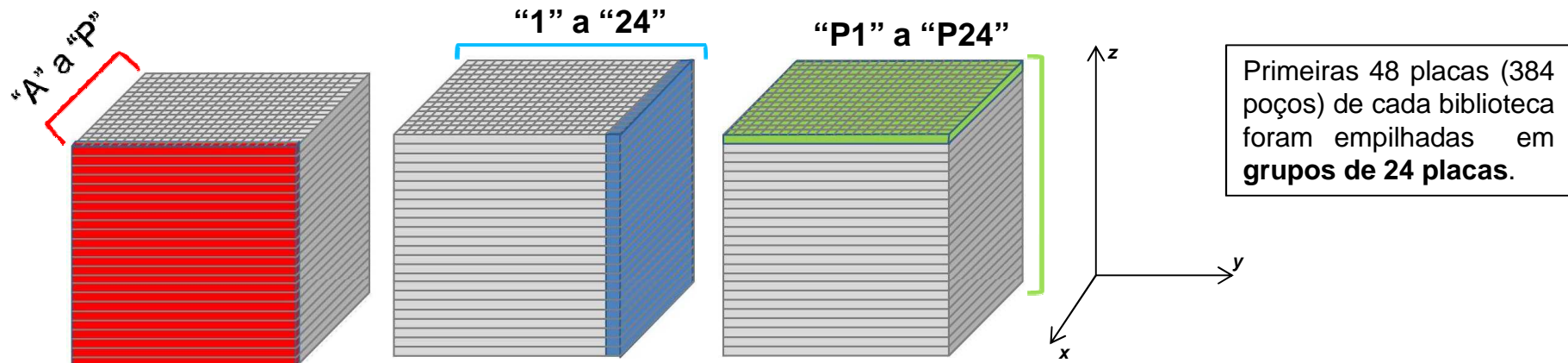


C1000™ Thermal Cycler



2. Identificação de clones BAC portadores de genes de interesse

a) Construção de pools 3D



SuperPool (conjunto de 24 placas):

- 24 *pools* plano X (colunas) ■
- 16 *pools* plano Y (linhas) ■
- 24 *pools* plano Z (placas) ■

Biblioteca ECG_Ba : SP1

	1	2	3	4	5	6	7	8	9	10	11	12
A			A	B	1	2	17	18	P9	P10		
B			C	D	3	4	19	20	P11	P12		
C			E	F	5	6	21	22	P13	P14		
D			G	H	7	8	23	24	P15	P16		
E			I	J	9	10	P1	P2	P17	P18		
F			K	L	11	12	P3	P4	P19	P20		
G			M	N	13	14	P5	P6	P21	P22		
H			O	P	15	16	P7	P8	P23	P24		

2. Identificação de clones BAC portadores de genes de interesse

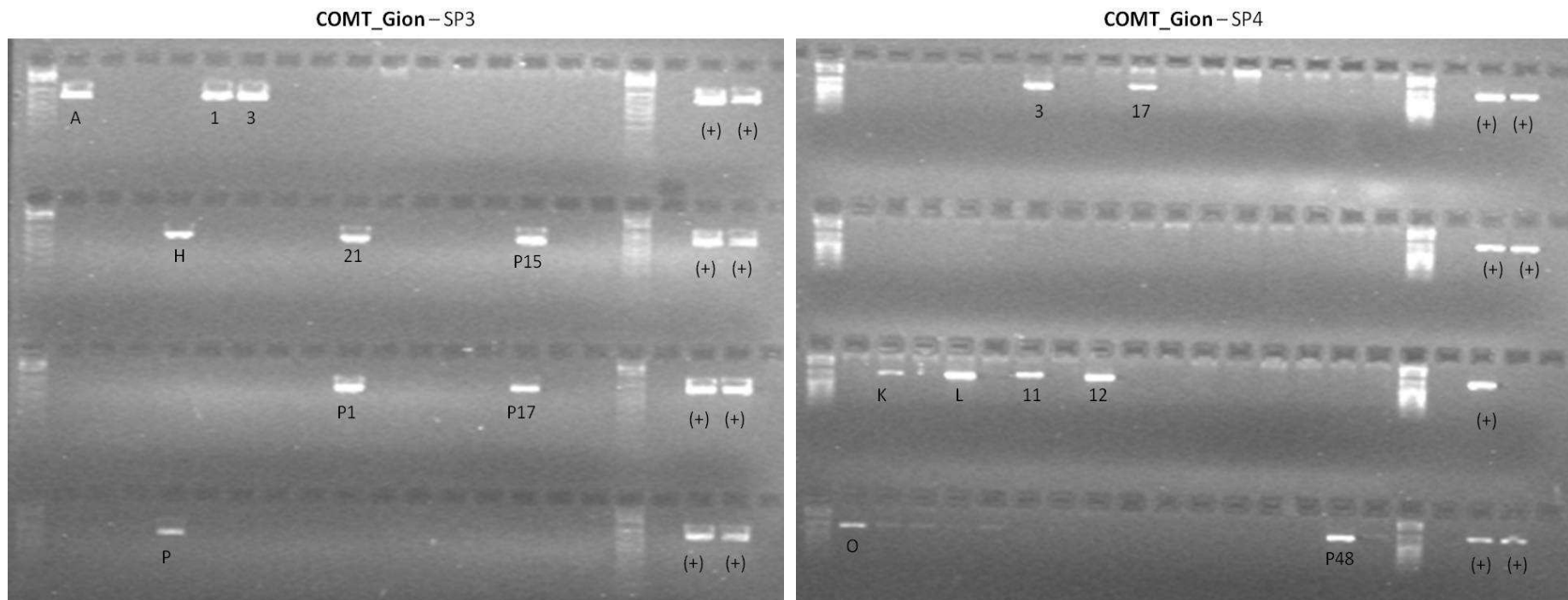
b) Identificação dos clones por PCR: *primers utilizados*

Família do gene	Gene	Acesso EMBL	Referência
Fenilalanina amonia-liase (PAL)	<i>EguPAL</i>	CT987001	Paiva et al, 2011
Cinamato 4-hidroxilase (C4H)	<i>EguC4H</i>	CT988030	Paiva et al, 2011
4-hidroxicinamoil – CoA ligase (4CL)	<i>Egu4CL</i>	AJ244010	Paiva et al, 2011
Hidroxicinamoil CoA: shikimato hidroxicinamoil transferase (HCT)	<i>EguHCT</i>	CT980202	Paiva et al, 2011
p-coumarato 3-hidroxilase (C3H)	<i>EguC3H</i>	CT986440	Paiva et al, 2011
Cafeoil CoA O-metiltransferase (CCoAOMT)	<i>EguCCoAOMT</i>	AF168778	Fernandes et al. 2009
Cinamoil-CoA redutase (CCR)	<i>EguCCR</i>	X79566	Rasmussen-Poblete et al. 2008
Ferulato 5 hidroxilase/ Coniferaldeído 5 hidroxilase (F5H/CAld5H)	<i>EguF5H</i>	CT987560	Paiva et al, 2011
Cafeato /5-hidroxiferulato O-metiltransferase (COMT)	<i>EguCOMT</i>	X74814	Fernandes et al. 2009
Cinamil álcool desidrogenase (CAD)	<i>EguCAD2</i>	X65631	Rasmussen-Poblete et al. 2008
GTPase RAC	<i>EguRAC1</i>	DR410036	Rengel et al. 2009
R2R3 factor de transcrição Myb	<i>EguMyb2</i>	AJ576023	Carver et al. 2005

2. Identificação de clones BAC portadores de genes de interesse

b) Identificação dos clones por PCR

- Identificação *pools* positivos



2. Identificação de clones BAC portadores de genes de interesse

b) Identificação dos clones por PCR

- Construção de diagrama com possíveis clones positivos

Coordenadas positivas para o gene COMT.

	SP1	SP2	SP3	SP4
x	-	5, 10	1, 3, 21	3, 11, 12, 17
y	-	G, H, M	A, H, P	K, L, O
z	-	P25, P34, P46	P1, P15, P17	P48



P48

K

L

O

Possíveis clones positivos

3

11

12

17

3

11

12

17

3

11

12

17

→ P48K03

→ P48K11

→ P48K12

→ P48K17

→ P48L03

→ P48L11

→ P48L12

→ P48L17

→ P48O03

→ P48O11

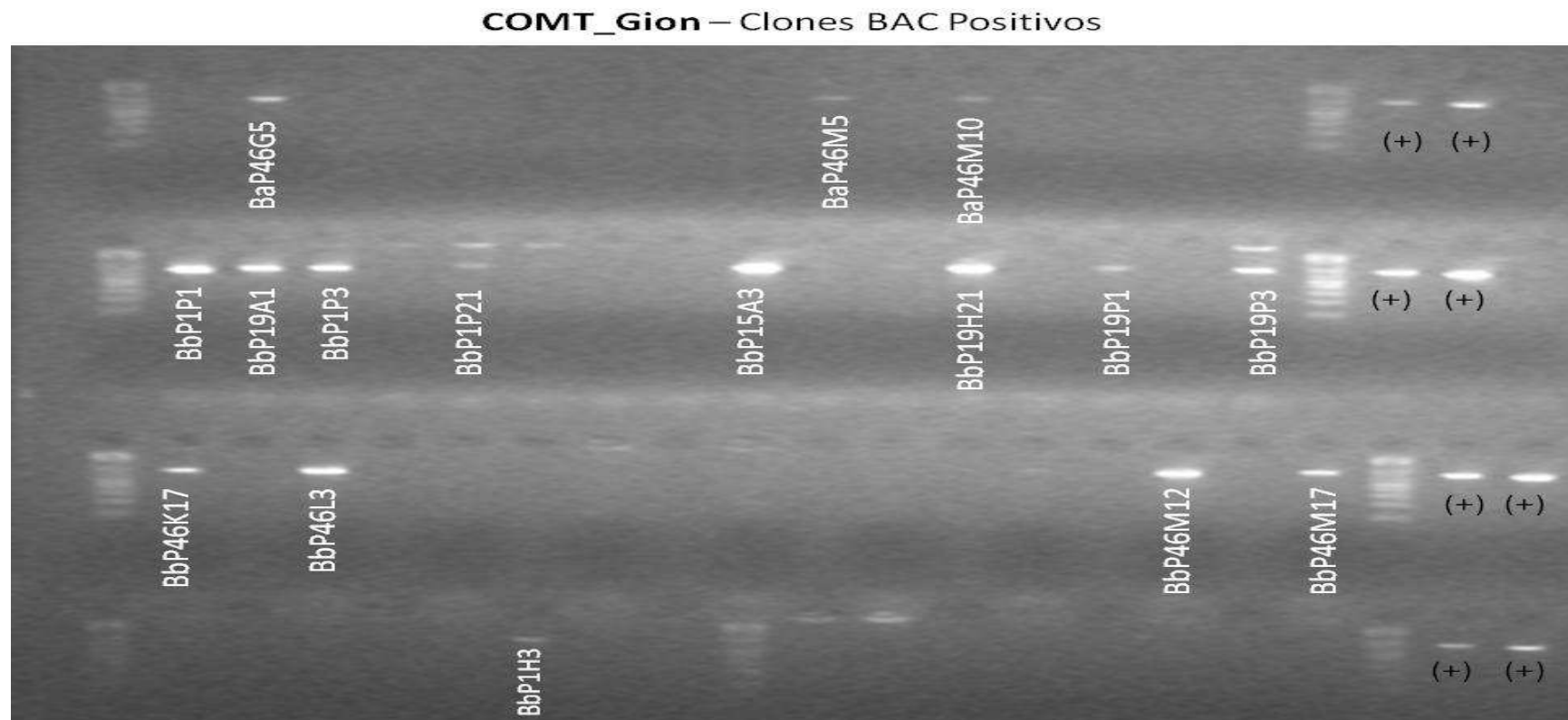
→ P48O12

→ P48O17

2. Identificação de clones BAC portadores de genes de interesse

b) Identificação dos clones por PCR

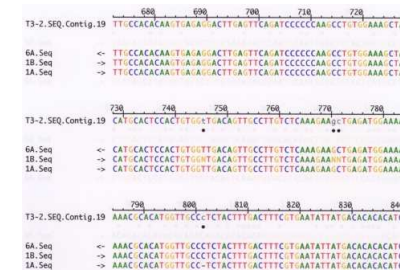
- Validação de clones positivos



2. Identificação de clones BAC portadores de genes de interesse

Resultados: Clones positivos por gene de interesse

Família do gene	Gene	EGC_Ba (a)	EGC_Bb (b)	Total (a+b)
Fenilalanina amonia-liase	<i>EguPAL</i>	3	3	6
Cinamato 4-hidroxilase (C4H)	<i>EguC4H</i>	0	0	0
4-hidroxicinamoil – CoA ligase	<i>Egu4CL</i>	0	1	1
Hidroxicinamoil CoA: shikimato hidroxicinamoil transferase	<i>EguHCT</i>	3	4	7
<i>p</i> -coumarato 3-hidroxilase	<i>EguC3H</i>	3	0	3
Cafeoil CoA O-metiltransferase	<i>EguCCoAOMT</i>	5	4	9
Cinamoil-CoA redutase	<i>EguCCR</i>	5	3	8
Ferulato 5 hidroxilase/ Coniferaldeído 5 hidroxilase	<i>EguF5H</i>	1	1	2
Cafeato /5-hidroxiferulato O-metiltransferase	<i>EguCOMT</i>	3	13	16
Cinamil álcool desidrogenase	<i>EguCAD2</i>	5	8	13
GTPase RAC	<i>EguRAC1</i>	6	1	7
R2R3 factor de transcrição Myb	<i>EguMyb2</i>	2	5	7
TOTAL		36	43	79



3. Caracterização de regiões do genoma nuclear

Resultados: Sequenciação e montagem

BAC	Primer	Leituras	% de leituras assembladas	Contigs > 500pb	Tamanho assemblado (pb)
EGC_Ba_P19K09	<i>EguPAL</i>	49.736	92	19	118.853
EGC_Ba_P32A22	<i>EguC3H</i>	72.618	93	24	143.159
EGC_Ba_P03C08	<i>EguCCoAOMT</i>	69.396	92	19	123.593
EGC_Ba_P26H12	<i>EguCCR</i>	39.929	92	21	228.151
EGC_Ba_P37H11	<i>EguCCR</i>	64.290	93	14	137.396
EGC_Ba_P15K22	<i>CCR_Rla</i>	62.486	92	17	109.635
EGC_Ba_P11O15	<i>CCR_Rlb</i>	62.877	95	13	182.734
EGC_Bb_P15A03	<i>EguCOMT</i>	36.791	91	7	94.741
EGC_Ba_P01L05	<i>EguCAD2</i>	42.480	93	5	131.716
EGC_Ba_P14L08	<i>CAD2_Rla</i>	27.736	92	8	168.755
EGC_Bb_P03K07	<i>EguMyb2</i>	34.382	92	9	116.644
EGC_Ba_P22O03	<i>EguRAC1</i>	80.593	93	37	138.702
EGC_Ba_P27H11	<i>EguRAC1</i>	39.689	91	11	98.741
Média		52.500	92		

Cobertura de sequenciação ≈ 133X

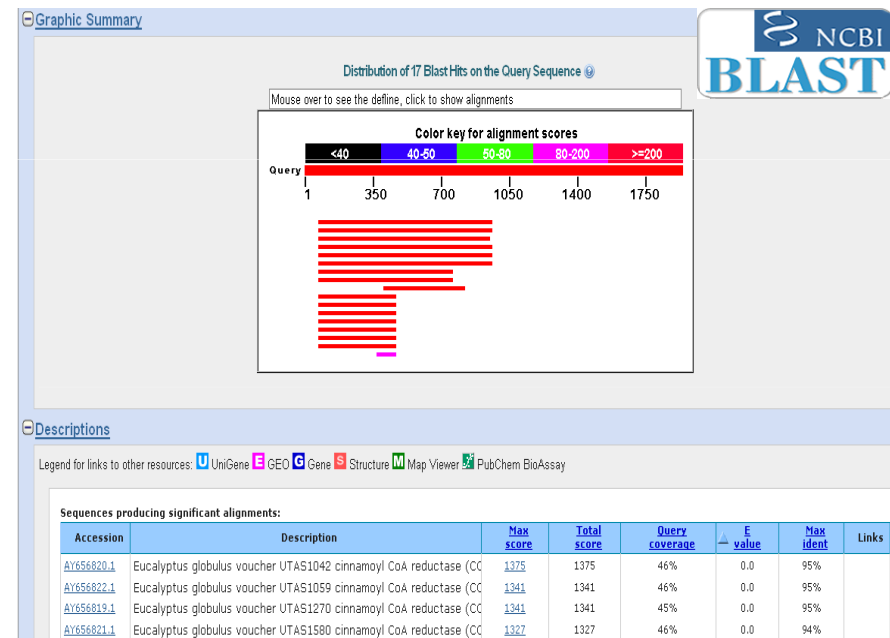
3. Caracterização de regiões do genoma nuclear

Análise de Sequências

Validação de clones BAC sequenciados

- Análise de *blast* identificou o gene de interesse para todos os clones BAC

- Identificação de outras sequências de interesse (CesA1 e CesA4)



BLAST® (<http://blast.ncbi.nlm.nih.gov/>)

3. Caracterização de regiões do genoma nuclear

Análise de Sequências

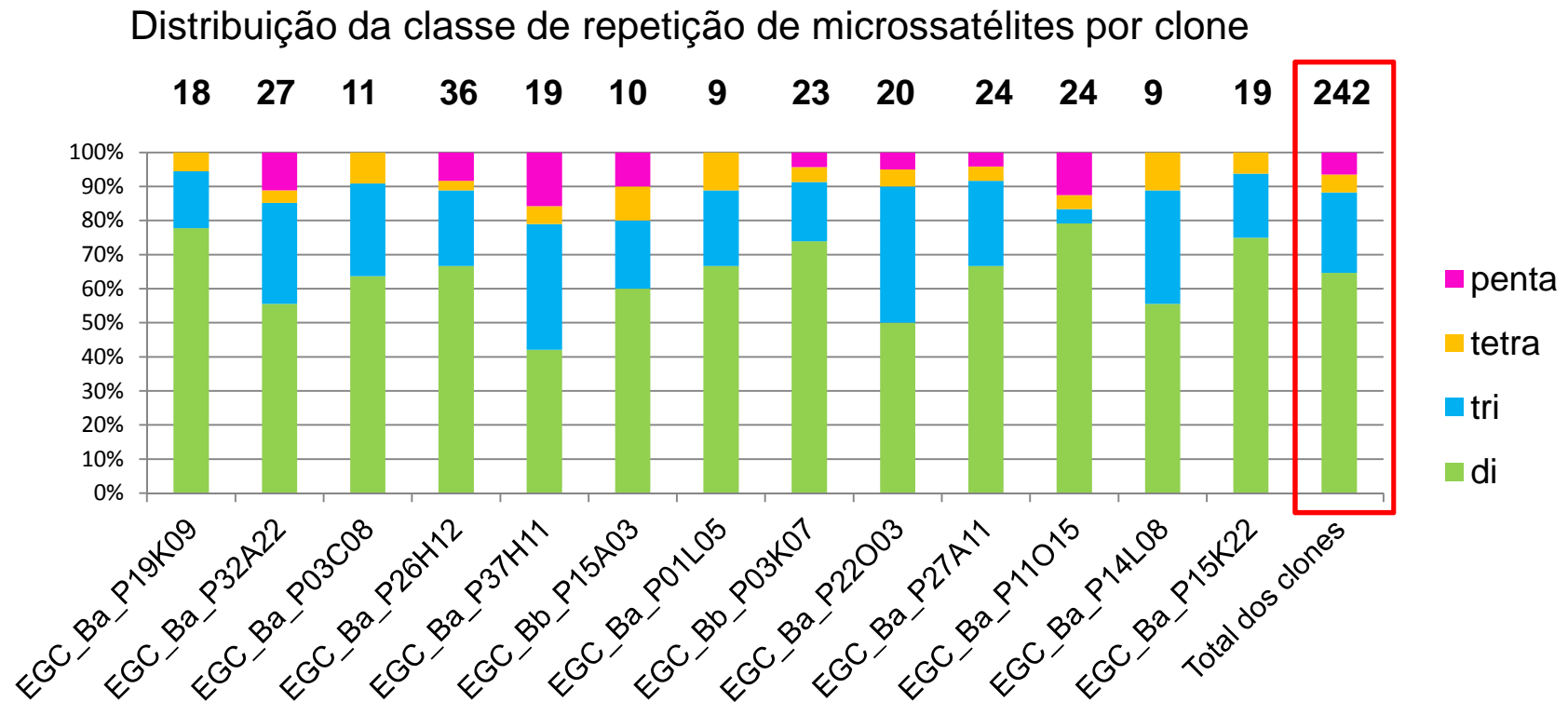
Elementos repetitivos

	Nº de elementos	Nº de bases (pb)	Percentagem do genoma (%)
Retroelementos	57	49374	2,75
<i>LINES</i> :	4	2050	0,11
L1 / CIN4	4	2050	0,11
<i>LTR</i> :	53	47324	2,64
Ty1 / Copia	38	37166	2,07
Gypsy / DIRS1	15	10158	0,57
ADN transponível	10	1394	0,08
hobo-Activator	3	742	0,04
MuDR-IS905	7	652	0,04
Total de repetições dispersas	67	50768	2,83
Repetições simples	397	15237	0,85
Baixa complexidade	58	2984	2,20

3. Caracterização de regiões do genoma nuclear

Análise de Sequências

Identificação de Microssatélites



Conclusão

- **As Bibliotecas BAC de *E.globulus*** apresentam:
 - cobertura de genoma bastante elevada (16X);
 - baixa taxa de contaminação organelar (3,33% EGC_Ba e 2,16% EGC_Bb);
 - probabilidade de encontrar uma sequência específica > 99,9%.
- Os ***pools* 3D** aumentam o rendimento do processo de identificação de clones de BAC via PCR ($\approx 85X$ menos reações PCR);

Conclusão

- Identificação de **clones positivos** para os genes de interesse:
 - 79 clones no total (COMT com 16 clones e C4H com 0 clones);
 - Valor médio 7-8 clones;
 - Cobertura esperada pelos *pools* para cada uma das bibliotecas.
- O recurso a diferentes **enzimas de restrição** aumenta a probabilidade de cobertura de todas as regiões presentes no genoma, bem evidente no caso da COMT.

Conclusão

- A **sequenciação** dos clones BAC (133X cobertura) permitiu:
 - Confirmar a presença dos genes de interesse nos clones seleccionados, bem como identificar outros genes;
 - caracterizar algumas das regiões do genoma relativamente à presença de **elementos repetitivos**:
 - **Elementos transponíveis** e de baixa complexidade com valores inferiores aos reportados para outras espécies de folhosas
 - **Microssatélites**, número elevado de repetições de dinucleótidos ((AG)_n) e de trinucleótidos ((AAG)_n, (AAT)_n e (AGG)_n)

Perspectivas futuras

- ✓ Desenvolver mapas físicos e sua integração nos mapas genéticos existentes;
- ✓ Estudar a microssintenia entre *E.globulus* e *E.grandis*;
- ✓ Determinar a riqueza de genes em regiões de interesse;
- ✓ Contribuir para a sequenciação integral do genoma de *E. globulus*.

Agradecimentos

- ❑ Professora Maria Helena Almeida (ISA-UTL)
- ❑ Doutor Jorge P. Paiva (IICT)
- ❑ Professor Pedro Fevereiro (BCV-ITQB)
- ❑ Doutora Susana Araújo (IICT)
- ❑ Drº Vitor Carocha (BCV-ITQB/IICT)
- ❑ Raiz – Instituto de Investigação da Floresta e do Papel
- ❑ CNRVG (Hélène Bergès, Sonia Vautrin e Elisa Prat)
- ❑ Universidade Paul Sabatier (Jacqueline Grima Pettenati)
- ❑ Fundação para a Ciência e Tecnologia (FCT)
 - Projectos:
 - ✓ GenEglob^{wq}
 - ✓ TREEFORJOULES
 - ✓ MicroEGo